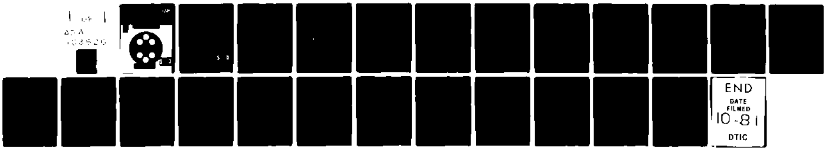
AD-A103 620   BROWN UNIV PROVIDENCE RI LEFSCHETZ CENTER FOR DYNAMI--ETC   F/G 12/1
AN AVERAGING METHOD FOR STOCHASTIC APPROXIMATIONS WITH DISCONTI--ETC(U)
MAY 81   H J KUSHNER

UNCLASSIFIED   LCDS-81-12   NL



LEVEL IV 12

Lefschetz Center for Dynamical Systems

Division of Applied Mathematics Brown University Providence RI 02912

AN AVERAGING METHOD FOR STOCHASTIC APPROXIMATIONS WITH DISCONTINUOUS DYNAMICS, CONSTRAINTS, AND STATE DEPENDENT NOISE

by

Harold J. Kushner

May 1981 LCDS Report 81-12

14 LCDS-81-12

6

AN AVERAGING METHOD FOR STOCHASTIC APPROXIMATIONS WITH DISCONTINUOUS DYNAMICS, CONSTRAINTS, AND STATE DEPENDENT NOISE[+]

10

Harold J. Kushner

Divisions of Applied Mathematics and Engineering
Lefschetz Center for Dynamical Systems
Brown University
Providence, Rhode Island 02912

11 May [redacted]81 12 23

[+]Work supported in part by the Air Force Office of Scientific Research under AFOSR 76-3063D, in part by the National Science Foundation under ~~NSF-Eng 77-12946-A02 and~~ in part by the Office of Naval Research under N00014-76-C-0279-[redacted].

15 AFOSR-76-3063

401834

BLANK PAGE

1. INTRODUCTION and ABSTRACT

Since the work of Robbins and Monro [1] and Kiefer and Wolfowitz [2], much attention has been devoted to the problems of stochastic approximation [3,4]. In recent years, work on the recursive estimation algorithms in adaptive control and communication systems has both rekindled widespread interest and required results under rather different assumptions on the noise and dynamics than were used in the early years (see, e.g., [5-10]). Problems with constraints have been treated by similar methods.

A typical form of current interest is the following: The iterate sequence is given by

$$(1.1) \qquad X_{n+1} = X_n + a_n h(X_n, \xi_n) \;, \quad X_n \in R^r \;, \quad \text{Euclidean } r\text{-space},$$

$\sum a_n = \infty$, $0 < a_n \to 0$ as $n \to \infty$, and $h(\cdot,\cdot)$ is not necessarily continuous. For example, it might be an indicator function. Also, the noise sequence $\{\xi_n\}$ might depend on $\{X_n\}$ in a complicated way. The references [5] - [10] contain a variety of techniques which are useful for proving w.p.1 or weak convergence for fairly general types of stochastic approximations, both with and without constraints. But they are not good enough to treat many problems where $h(\cdot,\cdot)$ is discontinuous or $\{\xi_n\}$ 'state dependent.'

In [11] and [12], averaging methods were used to get weak convergence results for suitably scaled stochastic difference equations, where the 'dynamical' term corresponding to our $h(\cdot,\cdot)$ might have the properties mentioned above. Here we adapt the method of [11] and [12], with that of [6] to develop a technique which is quite useful and versatile for the problems of interest. In a sense, we rely on the assumption that even if $h(\cdot,\cdot)$ is not smooth, expectations or conditional expectations of the types $Eh(\cdot,\xi_n)$ or $E[h(\cdot,\xi_n)|\xi_{n-1}, \xi_{n-2}, \ldots]$ are smooth functions of x. This situation occurs

in many cases, as attested to by the examples in Section 5. Results for both w.p.1 and weak convergence (see remark at end of Section 4) are available.

We also obtain analogous results for the following projection algorithm. Let $q_1(\cdot), \ldots, q_m(\cdot)$ denote continuously differentiable functions and define $G = \{x : q_i(x) \leq 0, \ i = 1, \ldots, m\}$. Let $\pi_G(y)$ denote any closest point in G to y. Then the algorithm is defined by

$$X_{n+1} = \pi_G\big(X_n + a_n h(X_n, \xi_n)\big) \quad . \tag{1.2}$$

In Section 2, we treat the algorithm (1.1), and the projected algorithm is treated in Section 3. A method for 'state dependent' noise $\{\xi_n\}$ is given in Section 4, and Section 5 contains two non-standard examples.

2. w.p.1 CONVERGENCE FOR (1.1)

Assumptions. Write δX_n for $X_{n+1} - X_n$, and let $\hat{C}_0^2$ denote the space of real valued continuous functions on R^r with compact support and continuous second partial derivatives. Let E_n denote expectation conditioned on $\{\xi_j, \ j < n\}$, and K will be used to denote a constant (its value might change from usage to usage).

One of the key difficulties is proving w.p.1 boundedness of $\{X_n\}$. For this we use a stability assumption on the differential equation $\dot{x} = \bar{h}(x)$, where $\bar{h}(x)$ is (*very loosely speaking*) $Eh(x,\xi_j)$. The boundedness argument uses a perturbed form of the Liapunov function $V(\cdot)$ for that differential equation, and various differences or derivatives of $V(\cdot)$ appear in the (mixing type) conditions. Owing to this, some of the conditions might seem at first glance a little unnatural, but they in fact are frequently readily verifiable. Theorem 1 and its conditions should be viewed as a prototype of a method which can be adapted to a wide variety of problems.

The assumptions are written such that (A1) - (A4) can be used for both bounded and unbounded $\{\xi_n\}$. With bounded $\{\xi_n\}$, we can let the $\alpha_{in} \equiv K$, all i,n. The (A6) and (A7) would not often hold as stated when $\{\xi_n\}$ is unbounded. The forms of (A6) and (A7) which are useful for the unbounded noise case depend on the particular form of $h(\cdot,\cdot)$ and it does not seem reasonable to try to get the most general form here. The unbounded case will be discussed after the theorem, and in the examples. Owing to our interest to have a proof which can be used with only minor changes when $\{\xi_n\}$ is unbounded, the details are a little more complicated than necessary.

A1. $\sum a_n^2 < \infty$, $\sum a_n = \infty$, $a_n > 0$, $\{a_{n+1}/a_n\}$ *is bounded.* $h(\cdot,\cdot)$ *is measurable and for each compact set* Q *there is a sequence* $\{\alpha_{0n}\}$ *such that* $|h(x,\xi_n)| \le \alpha_{0n}$ *for* $x \in Q$ *and* $\sum E\alpha_{0n}^2 a_n^2 < \infty$.

A2. *There is a twice continuously differentiable Liapunov function* $V(\cdot) \ge 0$ *such that* $|V_{xx}(\cdot)|$ *is bounded,* $V(x) \to \infty$ *as* $|x| \to \infty$. *There are* $\varepsilon_0 > 0$ *and* $\lambda_0 < \infty$ *such that for* $x \notin Q_0 \equiv \{x : V(x) \le \lambda_0\}$, *we have* $V_x'(x)\bar{h}(x) \le -\varepsilon_0$, *where* $\bar{h}(\cdot)$ *is defined in* (A3).

A3. *Let* $\{\alpha_{1n}\}$ *denote a sequence of random variables such that* $\sum_n E\alpha_{1n}^2 < \infty$. *There is a continuously differentiable* $\bar{h}(\cdot)$ *such that the limit defined (pointwise in* x) *by*

$$V_0(x,n) = \sum_{j=n}^{\infty} a_j V_x'(x) E_n[h(x,\xi_j) - \bar{h}(x)]$$

exists and (together with the partial sums) is bounded by $\alpha_{1n}(1 + |V_x'(x)\bar{h}(x)|)$. *The bound also holds if* $V(\cdot)$ *is replaced by a continuously differentiable function with compact support (*$\{\alpha_{in}\}$ *can depend on the function).*

A4. *There is a random sequence* $\{\alpha_{2n}\}$ *such that*

$$E_n|h(x,\xi_n)|^2 \le \alpha_{2n}^2(1 + |V_x'(x)\bar{h}(x)|)$$

and $a_n \alpha_{2n}^2 \to 0$ w.p.1 *as* $n \to \infty$.

A5. $|V_x'(x)\bar{h}(x)| \leq K(1 + V(x))$.

A6. *With* $[\ ]_x$ *denoting gradient with respect to* x, *let*

$$\left|\sum_{j=n}^{\infty} a_j [V_x'(x)\{E_n h(x,\xi_j) - \bar{h}(x)\}]_x\right| \leq K a_n \left(1 + |V_x'(x)\bar{h}(x)|^{1/2}\right)$$

The inequality holds with $V(\cdot)$ *replaced by an arbitrary continuously differentiable function with compact support.*

A7. *For* $s \leq 1$,

$$E_n |V_x'(x + sa_n h(x,\xi_n))\bar{h}(x + sa_n h(x,\xi_n))| \leq K(1 + |V_x'(x)\bar{h}(x)|)$$

Remark. As seen from the Examples in Section 5, the assumptions include some hard and interesting cases. In a sense, the 'prototype' model for (A1) - (A7) is the case where $|h(\cdot,\cdot)|$ and $|\bar{h}(\cdot)|$ have at most a linear growth in $|x|$ as $|x| \to \infty$ and $V(\cdot)$ has a growth one order higher than that of $h(\cdot,\cdot)$ and $\bar{h}(\cdot)$. Then the bounds in the assumptions make sense, under various mixing type conditions on $\{\xi_n\}$. Here, $\{\xi_n\}$ is not treated as being explicitly 'state dependent.' In the state dependent case, we must take into account the way that $\{\xi_n\}$ evolves as a function of $\{X_n\}$, and use a slightly different form of $V_0(x,n)$. See Section 4 and Example 2.

Theorem 1. Assume (A1) - (A7). The sequence $\{X_n\}$ is bounded w.p.1. If $V_x'(x)\bar{h}(x) \leq 0$ for all x, then $X_n \to \Lambda_0 \equiv \{x : V_x'(x)\bar{h}(x) \leq 0\}$ w.p.1. Otherwise, $\{X_n\}$ converges w.p.1 to the largest bounded invariant set $\mathcal{S}$[†] of

[†] Let S denote a bounded invariant set of (2.1). Then for each $x \in S$, there is a trajectory $x(\cdot)$ of (2.1) contained in $\mathcal{S}$ for $t \in (-\infty,\infty)$ and $x(0) = x$. The invariant set is the set of all limit points of bounded trajectories (on $[0,\infty)$) of (2.1).

$$(2.1) \qquad \dot{x} = \bar{h}(x) .$$

Remark. If $V_x'(x)\bar{h}(x)$ is not ≤ 0 for all x or if Λ_0 contains more than one point, then the limit set of $\{X_n\}$ might be more than one point, or even be a non-degenerate trajectory of (2.1). Such possibilities do exist in applications. But the theorem can be refined as follows. Let $x_0 \equiv x(t)$ be an asymptotically stable solution of (2.1) (in the sense of Liapunov) with domain of attraction $DA(x_0)$. There is a null set N such that if $\omega \notin N$ and $X_n(\omega) \in$ compact $A \subset DA(x_0)$ infinitely often, then $X_n(\omega) \to x_0$. The proof follows from the techniques of the proof below and the proof of Theorem 2.3.1 of [6].

Proof. The proof uses an 'averaged' form of the Liapunov function $V(\cdot)$. We have

$$(2.2) \qquad E_n V(X_{n+1}) - V(X_n) = a_n V_x'(X_n) E_n h(X_n,\xi_n) + a_n^2 \int_0^1 E_n h'(X_n,\xi_n) V_{xx}(X_n + s\delta X_n) h(X_n,\xi_n)(1-s)ds .$$

By (A2) and (A4), the last term is $\leq a_n \varepsilon_{1n}\left(1 + |V_x'(X_n)\bar{h}(X_n)|\right)$, where $\varepsilon_{1n} \to 0$ w.p.1, as $n \to \infty$. We also have

$$(2.3) \quad E_n V_0(X_{n+1},n+1) - V_0(X_n,n) =$$

$$E_n \sum_{n+1}^{\infty} a_j V_x'(X_{n+1}) E_{n+1}[h(X_{n+1},\xi_j) - \bar{h}(X_{n+1})]$$
$$- \sum_{n+1}^{\infty} a_j V_x'(X_n) E_n [h(X_n,\xi_j) - \bar{h}(X_n)]$$
$$- a_n V_x'(X_n)[E_n h(X_n,\xi_n) - \bar{h}(X_n)] ,$$

which we rewrite as

$$(2.4)\qquad -a_n V_x'(X_n)[E_n h(X_n,\xi_n) - \bar{h}(X_n)] + a_n E_n h'(X_n,\xi_n) \cdot$$

$$\int_0^1 ds \sum_{j=n+1}^{\infty} a_j [E_{n+1} V_x'(X_n+s\delta X_n)\big(h(X_n+s\delta X_n,\xi_j) - \bar{h}(X_n+s\delta X_n)\big)]_x \quad .$$

By (A6) and (A7) and an application of Schwarz's inequality, the last term of (2.4) is bounded by

$$(2.5)\qquad K a_n^2 \big(1 + |V_x'(X_n)\bar{h}(X_n)|\big) \quad .$$

Define the 'averaged' Liapunov function $\tilde{V}(n) = V(X_n) + V_0(X_n,n)$ and note that by (A3),

$$(2.6a)\qquad |V_0(X_n,n)| \le \alpha_{1n}\big(1 + |V_x'(X_n)\bar{h}(X_n)|\big) \quad ,$$

$$(2.6b)\qquad \tilde{V}(n) \ge -O(\alpha_{1n}) \quad ,$$

where $\alpha_{1n} \to 0$ w.p.1 as $n \to \infty$. Combining (2.2) - (2.5),

$$(2.7)\qquad E_n\tilde{V}(n+1) - \tilde{V}(n) = a_n(1+\delta_n)V_x'(X_n)\bar{h}(X_n) + \tilde{\delta}_n a_n \quad ,$$

where δ_n and $\tilde{\delta}_n$ go to zero w.p.1 as $n \to \infty$.

Define $\{m_n, M_n\}$ by

$$(2.8)\qquad \tilde{V}(n) - \sum_{i=0}^{n-1} a_i(1+\delta_i)V_x'(X_i)\bar{h}(X_i) - \sum_{i=0}^{n-1} \tilde{\delta}_i a_i = \sum_{i=0}^{n-1} m_i = M_n \quad .$$

By (2.7), $\{M_n\}$ is a martingale. By modifying $\{X_n, \delta_n, \tilde{\delta}_n, \xi_n\}$ on a set of arbitrarily small probability, we can suppose that there is an $N_0 < \infty$ such that $|\tilde{\delta}_i| \le \varepsilon_0/4$, $|\delta_i| \le 1/4$ for $i \ge N_0$. This modification will not alter the conclusions.

Let n_0 be a stopping time $\ge N_0$ and such that $X_{n_0} \notin Q_0$ (with n_0 equal to ∞ if $X_n \in Q_0$, all $n \ge N_0$). Define $n_1 = \min\{n : n>n_0,\ X_n \in Q_0\}$. Then by (2.6), (2.7) the sequence $\{\tilde{V}(n \cap n_1),\ n \ge n_0\}$ is a super martingale which is bounded below by $-O(\alpha_{1n})$. The facts that $E_n\tilde{V}(n+1) - \tilde{V}(n) \le -\varepsilon_0 a_n/2$

for $X_n \notin Q_0$ and n large, and $\sum a_n = \infty$ imply that $X_n \in Q_0$ infinitely often w.p.1. Define $Q_1 = \{x : V(x) \le \lambda_1\}$ where $\lambda_1 > \lambda_0$. Let Q_1 be the Q of (A1).

By a modification of the paths on a set of arbitrarily small measure, we can suppose that $|a_n h(x,\xi_n)| \le a_n \alpha_{0n} \le 1$ for large n and $x \in Q$ (say, for convenience, $n \ge N_0$). This modification will not affect the conclusions.

By (A1) and (A3), there are real $K_i(Q_1)$ such that if $X_n \in Q_1$,

$$|m_n|^2 \le \left\{|V_0(X_n,n)|^2 + |V_0(X_{n+1},n+1)|^2 + K_1(Q_1)a_n^2 + |V(X_{n+1}) - V(X_n)|^2\right\}K$$

$$\le K_2(Q_1)[a_n^2 + \alpha_{1n}^2 + \alpha_{1,n+1}^2 + a_n^2|h(X_n,\xi_n)|^2] \tag{2.9}$$

$$\le K_3(Q_1)[a_n^2 + \alpha_{1n}^2 + \alpha_{1,n+1}^2 + a_n^2\alpha_{0n}^2] \equiv K_3(Q_1)\beta_n \quad .$$

Let n_2 be any stopping time such that $X_{n_2} \in Q_1$. Let $n_3 = \min\{n : X_n \notin Q_1, n \ge n_2\}$. Then by (2.9) (note that the right side is summable over all return periods in Q_1),

$$P\{\sup_{n_2 \le n < n_3} |\sum_{i=n_2}^{n} m_i| \ge \varepsilon\} \le K_3(Q_1)E\sum_{i=n_2}^{n_3-1} \beta_i/\varepsilon^2 \quad . \tag{2.10}$$

We conclude from (2.8) and (2.10) and the recurrence of Q_0 and the facts that $a_n h(X_n,\xi_n) \to 0$ w.p.1 for $X_n \in Q_1$ and $V_x'(x)\bar{h}(x) \le -\varepsilon_0$ for $x \notin Q_0$ that eventually X_n remains in Q_1 for any $\lambda_1 > \lambda_0$. Hence M_n converges w.p.1.

Furthermore, since M_n converges w.p.1 and $V_0(X_n,n) = O(\alpha_{1n}) \to 0$ w.p.1 (since X_n eventually remains in the bounded set Q_1),

$$\sup_{m \ge n} |V(X_m) - V(X_n) - \sum_{i=n}^{m-1} a_i V_x'(X_i)\bar{h}(X_i)| \to 0 \tag{2.11}$$

w.p.1 as $n \to \infty$. If $V_x'(x)\bar{h}(x) \le 0$ all x, then (2.11) implies that $X_n \to \{x : V_x'(x)\bar{h}(x) = 0\}$.

Next, fix $f(\cdot) \in \hat{C}_0^2$, and repeat the development with $f(\cdot)$ replacing $V(\cdot)$. This yields

$$\sup_{m \geq n} \left| f(X_m) - f(X_n) - \sum_{i=n}^{m-1} a_i f_x'(X_i) \bar{h}(X_i) \right| \to 0$$

w.p.1 as $n \to \infty$. In particular, since for large n, $X_n \in Q_1$, we have

$$\sup_{m \geq n} \left| X_m - X_n - \sum_{i=n}^{m-1} a_i \bar{h}(X_i) \right| \to 0 \quad \text{w.p.1} \quad \text{as} \quad n \to \infty \,. \tag{2.12}$$

Equation (2.12) implies that (w.p.1) the limit points are contained in the set of limit points of the bounded trajectories of (2.1). Q.E.D.

<u>Remarks on the unbounded noise case</u>. (A6) and (A7), which were used only to get the bound (2.5), would not hold very often if $\{\xi_n\}$ were unbounded. If (2.5) were replaced by $a_n^2 \alpha_{3n} \left(1 + |V_x'(X_n) \bar{h}(X_n)|\right)$ where $\sup_n E a_n^2 \alpha_{3n}^2 < \infty$, and if this inequality holds with an arbitrary continuously differentiable $f(\cdot)$ with compact support replacing $V(\cdot)$, then the proof goes through with only minor changes. The $\{\alpha_{3n}\}$ can depend on $f(\cdot)$.

3. THE PROJECTION METHOD

Recall the definition of G and π_G from Section 1. Let $\bar{\pi}(\bar{h}(\cdot))$ denote the (not necessarily unique) projection of the vector field $\bar{h}(\cdot)$ onto G; i.e.,

$$\bar{\pi}(\bar{h}(x)) = \lim_{\Delta \downarrow 0} \left[\pi_G\left(x + \Delta \bar{h}(x)\right) - x\right] / \Delta \quad .$$

We will use

A8. *The* $q_i(\cdot)$, $i = 1, \ldots, m$, *are continuously differentiable,* G *is bounded, and is the closure of its interior* $G^0 = G - \partial G = \{x : q_i(x) < 0, \ i = 1, \ldots, m\}$.

In lieu of (A6), we use the weaker ('unbounded' noise) condition:

A9. *Let* $\{\alpha_{2n}\}$ *be* (see (A4)) *a random sequence satisfying* $E_n|h(x,\xi_n)|^2 \le \alpha_{2n}^2$ *for* $x \in G$. *Let* $\{\alpha_{4n}\}$ *be a random sequence satisfying* (*for* $x \in G$)

$$\left|\sum_{j=n}^{\infty} a_j[E_n h(x,\xi_j) - \bar{h}(x)]_x\right| \le \alpha_{4n}$$

and let $\alpha_{2n}E_n^{1/2}\alpha_{4,n+1}^2 \to 0$ *and* $\alpha_{2n}a_n \to 0$ w.p.1 *as* $n \to \infty$.

Remark. The assumption that G is the closure of its interior is useful in visualizing constructions in the proof, and slightly simplifies the details. The theorem remains valid when there are only equality constraints. Then, of course, $\{X_n\}$ moves on the constraint surface. If $\{\xi_j\}$ is bounded and satisfies a sufficiently strong mixing condition, then the α_{4n} in (A9) is $O(a_n)$ and the last requirement of (A9) holds by (A4). Condition (A9) is used only to show that (3.4) is of the order given below (3.4).

Theorem 2. Assume (A1), (A3) (with the V_x term dropped and for $x \in G$), (A8), (A9). Then (w.p.1) the limit points of $\{X_n\}$ are those of the 'projected' ODE

$$\dot{x} = \bar{\pi}(\bar{h}(x)) \quad . \tag{3.1}$$

Let $H(\cdot) \ge 0$ be a real valued function with continuous first and second partial derivatives and define $\bar{h}(\cdot) = -H_x(\cdot)$. Then, as $n \to \infty$, $\{X_n\}$ converges w.p.1 to the set $KT = \{x : \bar{h}'(x)\bar{\pi}(\bar{h}(x)) = 0\}$.

Remarks. The last remark after the statement of Theorem 1 also holds here. The form $\bar{h}(\cdot) = -H_x(\cdot)$ arises in the projected form of the Kiefer-Wolfowitz procedure, where we seek to minimize the regression $H(\cdot)$, subject to $x \in G$.

Proof. Except for the treatment of certain projection terms, the proof is quite similar to that of Theorem 1. Since G is bounded, the Liapunov function $V(\cdot)$ is not required and (A3, A6) and extensions need only be applied with $f(\cdot)$ an arbitrary real valued function with continuous second partial derivative, in order to characterize the limit points, similarly to what was done in Theorem 1.

We have

$$(3.2)\qquad E_n f(X_{n+1}) - f(X_n) = a_n f_x'(X_n) E_n h(X_n,\xi_n) + a_n f_x'(X_n) E_n \tau_n + a_n^2 \int_0^1 E_n(\delta X_n/a_n)' f_{xx}(X_n + s\delta X_n)(\delta X_n/a_n)(1-s)ds \quad ,$$

where τ_n is the 'projection error':

$$\tau_n = [\pi_G\big(X_n + a_n h(X_n,\xi_n)\big) - \big(X_n + a_n h(X_n,\xi_n)\big)]/a_n \quad .$$

If $X_n + a_n h(X_n,\xi_n) \notin G$, but there is a unique i such that $X_{n+1} \in \partial G_i = \{x : q_i(x) = 0\}$, then τ_n points 'inward' at X_{n+1} and, in fact, $\tau_n = -\lambda_i q_{i,x}(X_{n+1})$ for some $\lambda_i \geq 0$. In general, suppose that $X_n + a_n h(X_n,\xi_n) \notin G$, but (with a reordering of the indices, if necessary) X_{n+1} is in the intersection $\prod_1^\ell \partial G_i$. Then for each y for which $q_{i,x}'(X_{n+1})y \leq 0$, each $i \leq \ell$, we must have $\tau_n' y \geq 0$. (Otherwise τ_n would not be the 'projection error' or, equivalently, X_{n+1} would not be the closest point on ∂G to $X_n + a_n h(X_n,\xi_n) \notin G$.) Thus, by Farkas' Lemma, τ_n must lie in the cone $-C(X_{n+1})$, where

$$C(x) = \{y : y = \sum_{i\in A(x)} \lambda_i q_{i,x}(x)\} \quad ,$$

where $A(x)$ is the set of constraints which are active at x.

Note also that since $\delta X_n \to 0$ w.p.1 (by (A1)), there is a real sequence $0 < \mu_n \to 0$ such that $\tau_n = 0$ for large n (w.p.1) if distance $(X_n,\partial G) \geq \mu_n$,

and $a_n\tau_n$ and $a_n E_n \tau_n \to 0$ w.p.1. These facts will be used when characterizing the limit points below.

Now, following the argument in Theorem 1 but for smooth $f(\cdot)$ replacing $V(\cdot)$, define

$$f_0(x,n) = \sum_{j=n}^{\infty} a_j f_x'(x) E_n [h(x,\xi_j) - \bar{h}(x)]$$

and define $\tilde{f}(n) = f(X_n) + f_0(X_n,n)$. Analogous to the result in Theorem 1, there is a random sequence $\tilde{\delta}_n \to 0$ w.p.1 such that

$$E_n\tilde{f}(n+1) - \tilde{f}(n) - \tilde{\delta}_n a_n - a_n f_x'(X_n)\bar{h}(X_n) - a_n f_x'(X_n) E_n \tau_n = 0 . \tag{3.3}$$

We do not need to average out the $E_n\tau_n$ term. [In order to get (3.3), via the method of Theorem 1, we must show that the difference

$$E_n \sum_{j=n+1}^{\infty} f_x'(X_{n+1}) E_{n+1}[h(X_{n+1},\xi_j) - \bar{h}(X_{n+1})] - \sum_{j=n+1}^{\infty} f_x'(X_n) E_n [h(X_n,\xi_j) - \bar{h}(X_n)] \tag{3.4}$$

$$= E_n \delta X_n' \int_0^1 ds \sum_{j=n+1}^{\infty} a_j [f_x'(X_n+s\delta X_n) E_{n+1}\{h(X_n+s\delta X_n,\xi_j) - \bar{h}(X_n+s\delta X_n)\}]_x$$

is of an order $a_n\alpha_{5n}$ where $\alpha_{5n} \to 0$ w.p.1 as $n \to \infty$. But by (A1), (A9), and an application of Schwarz's inequality, we have $\alpha_{5n} = \alpha_{2n} E_n^{1/2} \alpha_{4,n+1}^2$.]

We also have

$$\tilde{f}(n) - \tilde{f}(0) - \sum_{i=0}^{n-1} a_i f_x'(X_i)\bar{h}(X_i) - \sum_{i=0}^{n-1} a_i f_x'(X_i)\tau_i - \sum_{i=0}^{n-1} a_i \bar{\delta}_i \equiv \sum_{i=0}^{n-1} m_i \equiv M_n , \tag{3.5}$$

where $\{M_n\}$ is a martingale and† $E\sum m_i^2 < \infty$. Finally, letting $f(\cdot)$ equal an arbitrary coordinate variable in G, and using the above square integrability and the fact that $f_0(X_n,n) \to 0$ w.p.1, we get

$$\sup_{m \geq n} \left| X_m - X_n - \sum_{i=n}^{m-1} a_i \bar{h}(X_i) - \sum_{i=n}^{m-1} a_i \tau_i \right| \to 0 \text{ w.p.1 as } n \to \infty . \tag{3.6}$$

---

†To get the inequality, we might have to alter $\{X_n,\xi_n\}$ on a set of arbitrarily small probability, but as in Theorem 1, this does not alter the conclusions.

By the properties of the 'projection terms' $\{a_n \tau_n\}$, and the fact that the 'limit dynamics' implied by (3.6) is that of the 'projected' ODE (3.1), (3.6) implies that (w.p.1) all limit points of $\{X_n\}$ must be limit points of (3.1). The $\sum a_i \tau_i$ term simply compensates for the part of $\sum a_i \bar{h}(X_i)$ which would take the trajectory out of G.

Now, let $\bar{h}(\cdot) = -H_x(\cdot)$, and use $H(\cdot)$ as a *Liapunov function*. Then

$$\dot{H}(x) = H_x(x)\bar{\pi}\big(-H_x(x)\big) \leq 0 \ . \tag{3.7}$$

Equation (3.9) implies that the limit points of (3.1) are contained in KT. Q.E.D.

4. STATE DEPENDENT NOISE

It is often necessary to take explicit account of the way that the evolution of $\{\xi_j,\ j \geq n\}$, depends on $\{X_j,\ j \leq n\}$. We might use a parametrization of the type $\xi_n = g_n(\xi_{n-1}, X_n, X_{n-1}, \ldots, X_{n-k}, \psi_n)$, where $\{\psi_n\}$ is an "exogenous' sequence. Such a scheme was used in [5] and [6], where the g_n were assumed to be sufficiently smooth functions of the $X_n, X_{n-1}, \ldots$. In the development of this section, we suppose that $\{X_n, \xi_{n-1},\ n \geq 1\}$ is a Markov process (not necessarily stationary). In fitting this format to particular applications, it might be required to 'Markovianize' the original (state, noise) process. Let E_n denote conditioning an $\{\xi_j,\ j < n,\ X_j,\ j \leq n\}$. Define the 'partial' transition function as follows. Define

$$P(\xi,n,\Gamma,n+1|x) = P\{\xi_{n+1} \in \Gamma | \xi_n = \xi,\ X_{n+1} = x\} \ .$$

In general, define $P(\xi,n,\Gamma,n+\alpha|x)$ by the convolution

$$P(\xi,n,\Gamma,n+\alpha+\beta|x) = \int P(\xi,n,dy,n+\alpha|x)P(y,n+\alpha,\Gamma,n+\alpha+\beta|x) \ .$$

Thus in calculating the above transition function, X_j is held fixed at x for $j \leq n+\alpha+\beta$. This partial transition function is useful because, loosely speaking, $\{\xi_n\}$ varies much faster than $\{X_n\}$ does. $V_0(x,n)$ is now written in the form

$$(4.1) \qquad V_0(x,n) = \sum_{j=n}^{\infty} a_j V_x'(x)\{\int h(x,\xi)P(\xi_{n-1},n-1,d\xi,j|x) - \bar{h}(x)\} .$$

Define $\tilde{V}(n) = V_0(X_n,n) + V(X_n)$. Note the way the averaging is done in (4.1) compared to how it was done in the $V_0(x,n)$ of (A3). The integral in (4.1) could be written as $E[h(x,\xi_j(x))|\xi_{n-1}(x) = \xi_{n-1}]$ where for each x,n, $\{\xi_j(x), j \geq n-1\}$ is a process which evolves according to the law $P(\xi,\alpha,\Gamma,\beta|x)$, where $\beta \geq \alpha \geq n-1$ and $\xi_{n-1}(x) = \xi_{n-1}$. See [12] for other applications of this idea.

Suppose that the sum in (4.1) is continuously differentiable in x, and that the derivatives can be taken termwise. Then (4.2) replaces the sum in (A6).

$$(4.2) \qquad \sum_{j=n}^{\infty} a_j [V_x'(x)\{\int h(x,\xi)P(\xi_{n-1},n-1,d\xi,j|x) - \bar{h}(x)\}]_x .$$

Theorem 3. <u>Assume</u> (A1) - (A7) <u>but with the above cited replacements</u> (4.1), (4.2). <u>Then the conclusions of Theorem</u> 1 <u>hold. The extensions to the unbounded noise case stated in the remark after Theorem</u> 1 <u>also hold here. Under the conditions of Theorem</u> 2, <u>subject to the above replacements, the conclusions of Theorem</u> 2 <u>hold</u>.

The proof is almost identical to that of Theorem 1 and (where appropriate) Theorem 2. We note only the following. By the Markov property,

$$(4.3) \qquad E_n P(\xi_n,n,\Gamma,j|X_n) = P(\xi_{n-1},n-1,\Gamma,j|X_n) , \quad j \geq n .$$

Note that the lowest term in (4.6) is (with $x = X_n$) $a_n V_x'(X_n)[E_n h(X_n,\xi_n) - \bar{h}(X_n)]$, exactly as in the sum in (A3). In the proof we get (4.4), the analog of the

first two terms on the right of (2.3).

$$(4.4)\quad \sum_{j=n+1}^{\infty} E_n a_j V_x'(X_{n+1})\{\int h(X_{n+1},\xi)P(\xi_n,n,d\xi,j|X_{n+1}) - \bar{h}(X_{n+1})\}$$

$$- \sum_{j=n+1}^{\infty} E_n a_j V_x'(X_n)\{\int h(X_n,\xi)P(\xi_{n-1},n-1,d\xi,j|X_n) - \bar{h}(X_n)\}$$

The left side of (4.3) replaces the right side of (4.3) in the second sum of (4.4). Then the differentiality assumption, and the bounds in (A1) - (A7), yield a bound analogous to the one obtained for the sums on the right of (2.3).

Remark. All the foregoing results hold if $\{a_n\}$ is random, under the following additional conditions. a_n depends on $\{X_i,\ i \le n\}$ only, $\sum_n a_n = \infty$, $\sum_n a_n^2 < \infty$, $\sum_n |a_{n+1} - a_n| < \infty$ w.p.1 and with

$$(4.4)\qquad \hat{V}_0(x,n) = a_n \sum_{j=n}^{\infty} E_n V_x'(x)[h(x,\xi_j) - \bar{h}(x)]$$

replacing the $V_0(x,n)$ of (A3).

Remarks on weak convergence. Define $t_n = \sum_0^{n-1} a_i$, $m_n = \min\{n : t_n \ge t\}$ and $X^0(t) = X_n$ on $[t_n,t_{n+1})$ and $X^n(t) = X(t+t_n)$ for $t \ge -t_n$ and $X^n(t) = X_0$ for $t \le -t_n$. Then the previous theorems imply various strong convergence properties for $\{X^n(\cdot)\}$. E.g., the proof of Theorem 1 implies that $\{X^n(\cdot)\}$ converges uniformly on finite time intervals to a solution of (2.1) and that the limit path is contained in the invariant set S cited there. Under weaker conditions, $\{X^n(\cdot)\}$ possesses various weak convergence (in $D^r[0,\infty)$) properties. Here, we only cite a result; the proof is quite similar to that of Theorem 8 in [15].

Assume the conditions of Theorem 1, except replace $\sum(1+E\alpha_{0n}^2)a_n^2 < \infty$ by $a_n \to 0$ and $\alpha_{0n}a_n \to 0$ w.p.1. Weaken (A3) to require only $\alpha_{1n} \to 0$ w.p.1 and $E\alpha_{1n}^2 \to 0$. Then $\{X^n(\cdot)\}$ is tight, and all limit paths are in S. Also $X_n \to S$ in probability. More strongly, for each $T < \infty$, $\epsilon > 0$,

$$\lim_{n\to\infty} P\{\sup_{|t|\le T} \text{distance}\,(X^n(t),S) \ge \epsilon\} = 0 .$$

If the conditions $Ea_n^2\alpha_{3n}^2 \to 0$ and $a_n\alpha_{3n} \to 0$ w.p.1 replace $\sum Ea_n^2\alpha_{3n}^2 < \infty$, then the above result also holds for the 'unbounded noise' case. There are similar weak convergence versions of Theorems 2 and 3.

5. EXAMPLES

5a. Example 1. Convergence of an adaptive quantizer. Frequently in telecommunications systems, the signal is quantized and only the quantized form is transmitted, in order to use the communications channel as efficiently as possible. It is desirable to adapt the quantizer to the particular signal [13,14], in order to maximize the quality of the received signal. Here a stochastic approximation form of an adaptive quantizer will be studied. Let $\xi(\cdot)$ denote the original stationary signal process and Δ a sampling interval. Write $\xi(n\Delta) = \xi_n$. The signal $\xi(\cdot)$ is sampled at instants $\{n\Delta,\ n = 0, 1, \ldots,\}$, a quantization $Q(\xi_n)$ calculated, and only this quantization is transmitted.

The quantizer is defined as follows. Let L denote an integer, x a parameter, and $\{\rho_i,\eta_i\}$ real numbers such that $0 = \rho_0 < \rho_1 < \ldots < \rho_{L-1} < \rho_L = \infty$, $0 = \eta_1 < \eta_2 < \ldots < \eta_L$. If $\xi_n > 0$, define $Q(\xi_n) = x\eta_i$ if $\xi_n \in [x\rho_{i-1}, x\rho_i)$, and set $Q(z) = Q(-z)$. In order to maintain the fidelity of the signal which is reconstructed from the sequence of received quantizations, the scaling parameter x should increase as the signal power increases.

Let $\beta \in (0,1]$ and let $0 < M_1 < \ldots < M_L < \infty$ with $M_1 < 1$, $M_L > 1$. A typical adaptive quantizer (adapting the scale x) is defined by (5.1), where X_n is the scale value at the *nth* sampling instant.

$$X_{n+1} = X_n^{\beta} B_n \quad , \quad B_n = M_i \quad \text{for} \quad X_n \rho_{i-1} \le |\xi_n| < X_n \rho_i \quad . \tag{5.1}$$

We will analyze a stochastic approximation version of (5.1). Let $\alpha > 0$ be such that $a_n \alpha < 1$, and let $\{\ell_i\}$ be real numbers such that $\ell_1 < 0$, $\ell_L > 0$, and $\ell_1 < \ell_2 < \ldots < \ell_L$. Let $0 < x_\ell < x_u < \infty$. Then we use

$$X_{n+1} = X_n^{(1-a_n\alpha)} (1 + a_n b_n) \Big|_{x_\ell}^{x_u} \tag{5.2}$$

where $b_n = \ell_i$ if $X_n \rho_{i-1} \le |\xi_n| < X_n \rho_i$, and the bar $|$ denotes truncation. With $\alpha > 0$, the algorithm has some desirable robustness properties. The algorithm (5.2) can be rewritten as (use $y^{1-\varepsilon} = y(1-\varepsilon \log y) + O(\varepsilon^2)$)

$$X_{n+1} = [X_n + a_n h(X_n, \xi_n) + O(a_n^2)] \Big|_{x_\ell}^{x_u} \quad , \tag{5.3}$$

where

$$h(X_n, \xi_n) = X_n b_n - \alpha X_n \log X_n$$

or

$$\begin{aligned} h(x,\xi) &= -\alpha x \log x + x \sum_{i=1}^{L} \ell_i I\{x\rho_{i-1} \le |\xi_n| < x\rho_i\} \quad . \\ \bar{h}(x) &= -\alpha x \log x + x \sum_{i=1}^{L} \ell_i P\{x\rho_{i-1} \le |\xi_n| < x\rho_i\} \quad . \end{aligned} \tag{5.4}$$

For specificity, let $\xi(\cdot)$ be a stationary Gaussian process. In particular for a matrix M whose eigenvalues have negative real parts and a standard Wiener process $w(\cdot)$, define $v(\cdot)$, $\xi(\cdot)$ by $dv = Mv\,dt + C dw$, $\xi = Dv$. Let $\sigma_0^2 = \text{var}\ \xi(t)$. Suppose that $\text{Cov}\ v(t) = \Sigma > 0$. We have

$$(5.5)\quad \frac{d}{dx}\left(\frac{\bar{h}(x)}{x}\right) = \frac{2}{\sqrt{2\pi}\,\sigma_0}\sum_{i=1}^{L} \ell_i\left[\rho_i \exp - \frac{\rho_i^2 x^2}{2\sigma_0^2} - \rho_{i-1}\exp - \frac{\rho_{i-1}^2 x^2}{2\sigma_0^2}\right] - \alpha/x$$

$$= \frac{2}{\sqrt{2\pi}\,\sigma_0}\sum_{i=1}^{L-1} (\ell_i - \ell_{i+1})\rho_i \exp - \rho_i^2 x^2/2\sigma_0^2 - \alpha/x \quad .$$

Thus $\bar{h}(x)/x$ is the sum of two strictly convex functions, the first being bounded and having a negative slope and the second going to ∞ as $x \to 0$ and to $-\infty$ as $x \to \infty$. Thus there is a unique $\bar{x} \in (0,\infty)$ such that $\bar{h}(\bar{x}) = 0$. Also $\bar{h}(x) > 0$ for $x < \bar{x}$ and $\bar{h}(x) < 0$ for $x > \bar{x}$. We use the 'unbounded' noise version of Theorem 2. See the remark after that theorem statement.

Let $\sum a_n^2 < \infty$. Since $h(\cdot,\cdot)$ is bounded (for $x \in [x_\ell, x_u]$), we need only verify (A3) for $f(\cdot) \in \hat{C}_0^2$ and (as noted in the remark after the proof of Theorem 1) get the appropriate bound for the second term of (2.4), with $f(\cdot) \in \hat{C}_0^2$ replacing $V(\cdot)$. Let E_n denote conditioning on $v(i\Delta)$, $i < n$.

It can be verified that (the rate of convergence of the sum depends on $v(n\Delta-\Delta)$)

$$(5.6)\qquad \sum_{j=n}^{\infty} E_n[h(x,\xi_j) - \bar{h}(x)] \le a_n K\big(|v(n\Delta-\Delta)| + 1\big) \quad .$$

The right-hand side of (5.6) goes to zero w.p.1 as $n \to \infty$ and (A3) holds. Next, using the fact that $(j \ge n)$ $P\{x\rho_{i-1} \le |\xi_j| < x\rho_i | v(n\Delta-\Delta)\}$ is a smooth and bounded function of x, and it and its x-derivative converge (fast enough) to the *unconditional probability* and its x-derivative as $j \to \infty$, we can get a bound of the form of the right-hand side of (5.6) on the last term of (2.4) (using $f(\cdot) \in \hat{C}_0^2$ in lieu of $V(\cdot)$). Thus all the conditions of the projection Theorem 2 hold. Hence if $\bar{x} \in [x_\ell, x_u]$, then $X_n \to \bar{x}$ w.p.1; otherwise X_n converges w.p.1 to the endpoint nearest to $\bar{x}$.

5b. <u>A Kiefer-Wolfowitz Procedure with Observation Averaging</u>. Let $f(\cdot)$ be a real valued function on R^1, whose first and second derivatives are bounded on R^1. Suppose that there is a unique θ (let $\theta = 0$) such that $f_x(\theta) = 0$ and let there be $\varepsilon_1, \varepsilon_2$ such that $f_x(x) \geq \varepsilon_1$ for $x \leq -\varepsilon_2$, and $f_x(x) \leq -\varepsilon_1$, for $x \geq \varepsilon_2$. Let $\{a_n, c_n\}$ be sequences of positive real numbers tending to zero and such that $\sum a_n = \infty$, $\sum a_n c_n < \infty$, $\sum a_n^2/c_n^4 < \infty$, $\limsup_n \sup_{j \geq n} a_j/a_n < \infty$, $\limsup_n \sup_{j \geq n} c_j/c_n < \infty$. Let $\{\psi_i\}$ be a sequence of mutually independent mean zero random variables with variances bounded by $\sigma^2 < \infty$, and *4th* moment by $m_4 < \infty$. Let $0 < \alpha < 1$, $\beta > 0$ and define $\{X_n, \xi_n\}$ by

$$X_{n+1} = X_n + a_n \xi_n , \quad n \geq 1 ,$$

$$\xi_n = \alpha\xi_{n-1} - \beta\left[\frac{f(X_n + c_n) - f(X_n - c_n)}{2c_n} + \frac{\psi_n}{c_n}\right] \tag{5.7}$$

$$= \alpha\xi_{n-1} - \beta f_y(X_n) + O_n - \beta\psi_n/c_n \quad ,$$

where $|O_n| \leq Kc_n$.

With $\alpha = 0$, we have a form of the Kiefer-Wolfowitz (KW) process. Then $\sum a_n^2/c_n^2 < \infty$ can replace $\sum a_n^2/c_n^4 < \infty$. With $0 < \alpha < 1$, the observations are averaged with exponentially decreasing weights. The conditions on $\{\psi_n\}$ and on $f(\cdot)$ can be relaxed, but the technique will be well illustrated with the given conditions. Here $h(x,\xi) = \xi$, which is not *a priori* bounded, and in fact is 'state' dependent. We show that $X_n \to 0$ w.p.1, via Theorem 1 (extension for unbounded noise). Define $\bar{h}(x) = -\beta f_x(x)/(1-\alpha)$. For notational convenience, <u>we drop the</u> O_n in (5.7). It can readily be carried through with little additional difficulty. Define $V(x) = x^2$.

Conditions (A2) and (A5) obviously hold. Define $\{\bar{\xi}_n, \hat{\xi}_n\}$ by $\bar{\xi}_n = \alpha\bar{\xi}_{n-1} - \beta f_x(X_n)$, $\hat{\xi}_n = \alpha\hat{\xi}_{n-1} - \beta\psi_n/c_n$. Clearly $\{\bar{\xi}_n\}$ is uniformly bounded and $\hat{\xi}_n = \beta \sum_1^n \alpha^{n-i} \psi_i/c_i$. Now,

$$a_n^2 E \hat{\xi}_n^2 \le \beta^2 \sigma^2 \sum_{i=1}^{n} \alpha^{2n-2i} a_n^2/c_i^2$$

and

$$\sum a_n^2 E \hat{\xi}_n^2 < \infty \quad .$$

Thus (A1) holds and $a_n \hat{\xi}_n \to 0$ w.p.1. Now check (A4). We have

$$(5.8)\quad E a_n^2 \Big(\sum_{i=1}^{n} \alpha^{n-i} \psi_i/c_i \Big)^4 \le a_n^2 \sum_{i,j=1}^{n} \alpha^{2(n-i)} \alpha^{2(n-j)} \sigma^4/c_i^2 c_j^2$$

$$+ a_n^2 \sum_{i=1}^{n} \alpha^{4(n-i)} m_4/c_i^4 \equiv \gamma_n \quad .$$

Since $\sum_n \gamma_n < \infty$, we have $a_n \hat{\xi}_n^2 \to 0$ w.p.1 and $a_n E_n \hat{\xi}_n^2 \to 0$ w.p.1. Thus (A4) holds.

Define $\{\bar{\xi}_j(x),\ j \ge n\}$, $\bar{V}_0(x,n)$, $\hat{V}_0(x,n)$, by

$$\bar{\xi}_j(x) = \alpha \bar{\xi}_{j-1}(x) - \beta f_x(x) \quad , \quad \bar{\xi}_n(X_n) = \bar{\xi}_n \quad ,$$

$$\bar{V}_0(x,n) = \sum_{j=n}^{\infty} a_j V_x(x) [\bar{\xi}_j(x) - \bar{h}(x)] \quad ,$$

$$\hat{V}(x,n) = \sum_{j=n}^{\infty} a_j V_x(x) E_n \hat{\xi}_j \quad .$$

Then $|\bar{V}_0(x,n)| \le K a_n (1+|x|)$. Also

$$E_n \hat{\xi}_j = \alpha^{j+1-n} \hat{\xi}_{n-1}$$

and

$$\hat{V}(x,n) = \Big(\sum_{j=n}^{\infty} a_j \alpha^{j+1-n} V_x(x) \Big) \xi_{n-1} \quad .$$

These representations can be used to readily show that both (A3) and the condition mentioned for the unbounded noise case after the proof of Theorem 1 hold. Thus, by (the unbounded noise extension of) Theorem 1, $X_n \to 0$ w.p.1.

REFERENCES

1. H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Statist.*, 22, 1951, 400-407.

2. J. Kiefer and J. Wolfowitz, "Stochastic estimation of the maximum of a regression function," *Ann. Math. Statist.*, 23, 1952, 462-466.

3. M. T. Wasan, *Stochastic Approximation* (Cambridge: Cambridge University Press, 1969).

4. M. B. Nevelson and R. Z. Khazminskii, *Stochastic Approximation and Recursive Estimation*, 1972; English Translation published by the American Mathematical Society, 1976.

5. L. Ljung, "Analysis of recursive stochastic algorithms," *IEEE Trans. on Automatic Control*, AC-22, 1977, pp. 551-575.

6. H. J. Kushner and D. S. Clark, *Stochastic Approximation Methods for Constrained and Unconstrained Systems*, Applied Math Science Series #26 (Berlin: Springer, 1978).

7. D. C. Farden, "Stochastic approximation with correlated data," *IEEE Trans. on Information Theory*, IT-27, 1981, pp. 105-113.

8. V. Solo, "The convergence of AML," *IEEE Trans. on Automatic Control*, AC-24, 1979, pp. 958-962.

9. A. Benveniste, M. Goursat, and G. Ruget, "Analysis of stochastic approximation schemes with discontinuous and dependent forcing terms with applications to data communication," *IEEE Trans. on Automatic Control*, AC-25, 1980, pp. 1042-1057.

10. H. J. Kushner, "A projected stochastic approximation method for adaptive filters and identifiers," *IEEE Trans. on Automatic Control*, AC-25, 1980, pp. 836-838.

11. H. J. Kushner and Hai Huang, "On the weak convergence of a sequence of general stochastic difference equations to a diffusion," *SIAM J. on Applied Math.*, 40, 1981, pp. 528-541.

12. H. J. Kushner and Hai Huang, "Averaging methods for the asymptotic analysis of learning and adaptive systems, with small adjustment rate," to appear *SIAM J. on Control and Optimization*.

13. D. J. Goodman and A. Gersho, "Theory of an adaptive quantizer," *IEEE Trans. on Communication*, COM-22, 1974, 1037-1045.

14. D. Mitra, "A generalized adaptive quantization system with a new reconstruction method for noisy transmission," *IEEE Trans. on Communication*, COM-27, 1979, 1681-1689.

15. H. J. Kushner, "Stochastic approximation with discontinuous dynamics and state dependent noise; w.p.1 and weak convergence," *Proceedings of the Conference on Decision and Control*, Albuquerque, 1980.

jbf